Back to the basics: Abstract painting as an index of creativity

Lucas Bellaiche, Anna P. Smith, Nathaniel Barr, Alexander Christensen, Chloe Williams, Anya Ragnhildstveit, Jonathan Schooler, Roger Beaty, Anjan Chatterjee & Paul Seli

Published online: 08 Aug 2023.

Submit your article to this journal ⇗

View related articles ⇗

View Crossmark data ⇗

Routledge
Taylor & Francis Group

Check for updates

Back to the basics: Abstract painting as an index of creativity

Lucas Bellaiche [ID][a][1], Anna P. Smith [ID][a][1], Nathaniel Barr[b], Alexander Christensen[c,d], Chloe Williams[a], Anya Ragnhildstveit[e], Jonathan Schooler[f], Roger Beaty[g], Anjan Chatterjee[c], and Paul Seli[a]

[a]Duke University; [b]Sheridan College; [c]University of Pennsylvania Perelman School of Medicine; [d]Vanderbilt University; [e]University of Cambridge; [f]University of California; [g]Pennsylvania State University

ABSTRACT

Researchers have invested a great deal in creating reliable, "gold-standard" creativity assessments that can be administered in controlled laboratory settings, though these efforts have come at the cost of not using ecologically and face-valid tasks. To help fill this critical gap, we developed and implemented a novel, face-valid paradigm that required participants to paint abstract pieces of art, which were later rated for creative quality. We first sought to evaluate whether there was good convergence among creativity ratings provided by independent raters. Next, we examined whether its measure of creativity correlated with (a) existing creativity measures and (b) individual traits (e.g. openness, fluid intelligence) that are typically correlated with indices of creativity. Our findings indicate that our abstract-painting paradigm is feasible to implement (independent ratings of the creativity of the paintings converged well), and that its measure of creativity significantly correlated with some of the gold-standard indices of creativity (thereby providing convergent validity). These findings suggest that having participants engage in abstract painting provides a valid index of creativity, thereby opening new opportunities for future research to index a more-face-valid measure of creativity.

Despite the longstanding centrality of creativity in human society, improving the understanding and measurement of creative ability via valid paradigms continues to be a primary challenge for psychological researchers (Ausubel, 1967; Guilford, 1950). This challenge has resulted in the development of many useful tools, including the Torrance Tests of Creative Thinking (Torrance, 1966), which assesses verbal and figural creativity, and the Alternate Uses Task (AUT; Guilford, 1967), which measures semantically divergent thinking. While these tasks have been thoroughly validated, show high inter-rater reliability, and are easy to implement, once concern is that they "immediately bump head-on to the nature of the creative act, which commonly is quite complex and to be recognized must have an observer capable of embracing its complexity" (Barron, 1965, p. 13). Put differently, these gold-standard paradigms have fallen short in their ability to capture and assess the complex and applied creative process in an ecological manner. Thus, the primary aim of this exploratory investigation was to develop a novel, more-face-valid creativity task that can be implemented in future research.

The need for ecologically based approaches toward creative ability is underscored by what current measures of creativity actually target. For instance, the AUT positions itself as a test of divergent-thinking ability. Although divergent thinking is not strictly synonymous with creativity, researchers often characterize it as an acceptable indicator of creative ability, based on its correspondence to real-world creative activities and self-reported creative achievements (Jauk, Benedek, & Neubauer, 2014; Runco & Acar, 2012). Yet, for all of the predictive power of AUT, one elephant remains in the room: the test does not resemble everyday instances of creativity, such as culinary experimentation, landscaping, web design, or songwriting, to use examples from the Creative Achievement Questionnaire (Carson, Peterson, & Higgins, 2005).

One possibility is that the AUT, and other laboratory measures of divergent thinking, measure the same, or a closely overlapping construct as do intelligence tests. Consistent with this view, much research has investigated the relationship between divergent-thinking tasks of creativity and general intelligence, finding that scores on general and fluid intelligence tests account for between 24 and 63% of variance on creativity,

CONTACT Lucas Bellaiche ✉ lucas.bellaiche@duke.edu 🖻 Department of Psychology & Neuroscience, Duke University, 417 Chapel Drive, Durham, NC 27708, USA

[1]These authors contributed equally to this work.

depending on how they are measured (Frith et al., 2021; Silvia & Beaty, 2012). A notable qualification to these findings is that Jauk, Benedek, Dunst, and Neubauer (2013) found that an individual's general intelligence could predict creative ability (i.e., ideational originality) only when intelligence falls below a certain threshold (i.e., creative ability is correlated with intelligence only among individuals with relatively low IQ scores). Critically, however, no such threshold of intelligence emerged for people's self-reported creative achievements: intelligence was found to predict creative achievement equally well across a wide range of IQ scores. What this suggests is that, whereas creative ability may be measuring something more akin to intelligence (at least among those with relatively low IQ scores), reports of creative achievement do not vary as a function of IQ.

However, Karwowski et al. (2016) further qualify this claim in a large analysis of eight studies examining correlational data between intelligence tests (e.g., Raven Matrices) and various creativity tasks (including divergent-thinking tasks, like the AUT, and self-report creative-achievement batteries). Using Necessary Condition Analyses as an alternative to the typical regression analyses used when testing the aforementioned "threshold hypothesis," they found support that "intelligence is a necessary-but-not-sufficient condition of creativity" (p. 107), even for the self-report measures of creative achievement. Thus, at odds with Jauk, Benedek, Dunst, and Neubauer (2013), Karwowski et al. suggest that creativity – as measured by self-report – may be influenced by intelligence, though the authors note the possibility of important differences between self-report versus "more objective" measures of observed creative behavior, and highlight the need for more research in this sphere (p. 115).

Overall, though self-report creative achievement has mixed findings in regards to its overlap with intelligence, research consistently suggests that popular task-based measures of creativity, such as the AUT, may inadvertently provide a substantial measure of IQ, and relatively little in the way of more ecologically valid creativity, per se. Thus, in line with Guilford's original call to develop measures that disentangle creativity from intelligence (Guilford, 1950), there is room for the development of novel measures of creative ability that deconfound measures of creativity and intelligence, and focus on more-applied and face-valid creativity tasks. And, despite decades of research, little is known about how similarly individuals perform on divergent-thinking tasks in relation to such externally valid tasks.

Notably, some research has employed tasks of creative ability that are more ecologically valid and better resemble instances of applied creative achievement than the common divergent-thinking tasks, although such research is relatively sparse. However, many of these procedures are, like the AUT, limited to an individual's verbal and written abilities, including creative humor production (Christensen, Silvia, Nusbaum, & Beaty, 2018; Nusbaum, Silvia, & Beaty, 2017), metaphor production (Beaty & Silvia, 2013; Christensen & Guilford, 1963; Silvia & Beaty, 2012), and writing (Johnson et al., 2022).

Outside of the verbal domain, some researchers have examined musical improvisation – from jazz pianists to classical musicians and freestyle rappers – to assess domains of non-verbal creativity (Beaty, 2015). More commonly, researchers have used drawing tasks such as the Test for Creative Thinking-Drawing Production (TCT-DP; Jellen & Urban, 1989), the Test of Creative Imagery Abilities (TCIA; Jankowska & Karwowski, 2015), and the "Figural" subset of the Torrance Tests of Creative Thinking (TTCT; Torrance, 1966). The TCT-DP provides participants with an unfinished drawing in the form of figural fragments, with which they must complete a cohesive drawing (which is then rated, by independent raters, across 14 criteria; Jellen & Urban, 1989). Similarly, the TCIA provides participants with an unfinished drawing, then prompts them to list as many ideas as possible for a new creative image that is based on the unfinished drawing (see Method, Appendix A). Next, participants select what they believe to be their most original idea, and then draw, elaborate on, and caption it (drawings are graded across three criteria of creative imagery; Jankowska & Karwowski, 2015). Lastly, the TTCT-Figural measures divergent thinking and other imagery features (e.g., emotional expressiveness), and consists of three tasks: (a) creatively elaborate on a basic picture, like a pear-shaped figure ("Construction"), (b) synthesize different incomplete figures into a cohesive image ("Completion"), and (c) utilize either lines or circles to make as many different images as possible within a time limit ("Lines/Circles") (Torrance, 1966; see Alabbasi, Paek, Kim, & Cramond, 2022).

Although, by approaching tests of creative potential with more ecologically valid instances of creativity, such tasks represent a step forward in the measurement validity of creative thinking, they are somewhat limited in that they do not allow full freedom of creative expression. Indeed, these drawing tasks constrain participants to use only two hues (gray of the pencil and white of the page), and they require participants to create creative figures by expanding upon/incorporating in their figures

shapes that are provided by the researcher. Thus, some of these more ecologically valid creativity tasks are reduced to the verbal domain (e.g., metaphor production), or, in the case of drawing tasks (e.g., TCT-DP, TCIA, TTCT), the requirements are relatively simple and constrained.

A progression toward incorporating externally valid creativity measures is Teresa Amabile's Consensual Assessment Technique (CAT; Amabile, 1982). In the CAT, researchers rely on the subjective ratings of creativity, as provided by expert raters. That is, raters of a creative product do not use a standardized definition of "creativity," but rather a subjective sense of what is "creative" to them. Such ratings, then, are inherently "relative and bounded by time and place" (Hennessey, Amabile, & Mueller, 2011, p. 255). Still, in the CAT rating procedure, raters must agree on what is "creative" and be able to differentiate less- from more-creative products; put differently, there is a requirement for inter-rater reliability, which ensures that the experts' independent, subjective ratings of creativity tend to align. Importantly, this, in turn, provides some evidence of objectivity among the subjective ratings.

One important consideration of Amabile's method is the reliance on expert raters. While this definition varies across implementations of the procedure, the general consensus is that raters are considered "expert" if they have sufficient familiarity with the domain of the product; in specialized domains like abstract painting, there are thus limitations in the efficacy of using an expert population of raters given how rare expert-level knowledge could be in certain fields (Hennessey, Amabile, & Mueller, 2011). While the CAT relies on an expert view of creativity, our abstract-painting task extends work by Amabile by allowing non-expert raters (who are easier to come by) to consider the face-value creativity of a specific kind of visual art. Importantly, Hawley-Dolan and Winner (2011) found that both expert and non-expert viewers preferred and judged abstract artworks in similar manners, and that their findings support that "the world of abstract art is more accessible than people realize" (p. 435). Thus, by using non-experts to provide their subjective assessments of creativity, we hope to encourage external validity in an evaluation process akin to, for instance, non-artists appreciating work in a museum.

The present study

In developing a face-valid creative ability test that captures ecologically valid processes, we chose abstract painting as our exploratory creative ability paradigm; this decision was made for three primary reasons. First, abstract painting is, at face value, an inherently creative

process: Whereas one might reasonably question the extent to which completing the AUT, for example, reflects "creativity," it is widely agreed that abstract paintings involve a creative process. Corroborating this view, as noted by Halasz (2009), abstract painting requires a synthesis of ideas common to any creative expression.

Second, during abstract painting, participants are able to rely on unprobed inspiration to freely produce their work of creativity: There is inherent freedom in the creative process as participants are literally presented a blank slate upon which they can express their inner thoughts in an external medium.

Third, while certain creative tasks (e.g., jazz improvisation) require participants to have developed – over a protracted period of time – the skills necessary to perform the task (e.g., knowing how to play piano), abstract paintings can be completed by both amateurs and experts alike, and previous experience is not necessarily a prerequisite for participation in studies on abstract painting. Though Hawley-Dolan and Winner (2011) did find that participants could distinguish between abstract art created by children and by professional artists (implicating expertise of artists in judgments of their paintings to a degree), abstract paintings can nonetheless be created without prior technical training as compared to other forms of creative arts, and thus, may differentiate better at lower skill levels.

In light of the foregoing, in the present study we brought participants into the laboratory to freely paint abstract pieces of art. These abstract paintings were then rated – by independent raters – in terms of their creativity. To validate this task, we first sought to determine the feasibility of its implementation. Specifically, we examined whether (a) participants could provide paintings whose creativity ratings showed a reasonable amount of variability, and (b) creativity ratings, provided by independent raters, showed good reliability. Of additional interest, we sought to examine the domain-specificity (or lack thereof) of abstract painting, and did so by assessing whether either or both the AUT (Guilford, 1967) and the TCIA (Jankowska & Karwowski, 2015) correlated with creative quality of paintings. Finally, we investigated whether creative quality of paintings was associated with various common indices in creativity research, including personality and mindset measures (see Method for a full list of measures and their respective associations with creativity). This was done to explore if and how the creative quality of an individual's abstract painting overlapped with other indices of creativity facets. Given the exploratory nature of this work, we did not determine any specific a priori

hypotheses on the second or third aim, but instead intended to motivate future research that could directly assess creative tasks like abstract painting as a predictor for an individual's creative potential and achievement.

Method

Participants

This study was approved by the Institutional Review Board of Duke University. 100 participants were recruited from an undergraduate online subject pool and were offered course credit in return for one-hour participation in this study. Individuals were eligible if they were over the age of 18 and spoke English fluently. The data from one participant were excluded from analysis because they did not follow directions (i.e., they drew a representational, not abstract, painting), reducing the sample to a final sample size to 99 (mean age = 19.13, SD = 1.08, 31% male).

Participants were told at sign-up that they would be given a number of questionnaires that would assess their personality, fluid intelligence, creative mindset, and creative history. They were also told they would also complete three creativity tasks wherein they would be asked to perform activities such as completing a picture with the shape given or devising novel uses for everyday objects. Finally, participants were informed they would be asked to complete an abstract painting using the materials provided. They were told that the study would take approximately one hour, and that they would receive one course credit at completion.

Materials

Instrument text/instructions are included in Appendix A.

Primary creativity measures of interest

We administered three tasks to measure in-lab creative performance: a classic laboratory-based task in the verbal domain (AUT; Guilford, 1967), a laboratory task in the visuospatial domain (TCIA; Jankowska & Karwowski, 2015), and an ecologically valid opportunity to showcase visual artistic ability (abstract painting). To estimate the quality of responses on these measures, we used many-facet Rasch models (MFRM; Linacre, 1994) on the ratings of three trained experimenters.

Recent work using creativity tasks has demonstrated that MFRM are effective for correcting rater biases (e.g., being more severe or lenient than other raters; see Christensen, Silvia, Nusbaum, & Beaty, 2018; Primi, Silvia, Jauk, & Benedek, 2019; Silvia, Christensen, & Cotter, 2021 for implementations of MFRM in creativity tasks). Like single-facet Rasch models for ability tests, many-facet Rasch models adjust ability estimates for the lenience (or severity) of the raters but also the difficulty of the prompt (e.g., "pen" may be more difficult to come up with alternatives uses for than a "box"; Eckes, 2011). In our implementation, each person's estimated ability was adjusted for variation in the differences in the raters' severity as well as prompt difficulty for the AUT and TCIA.

Abstract painting task. Participants were given five paintbrushes, an assortment of different colored acrylic paints, six palette knives, one blank 11 in. x 14 in. canvas board, and one apron (see Figure 1). The participants
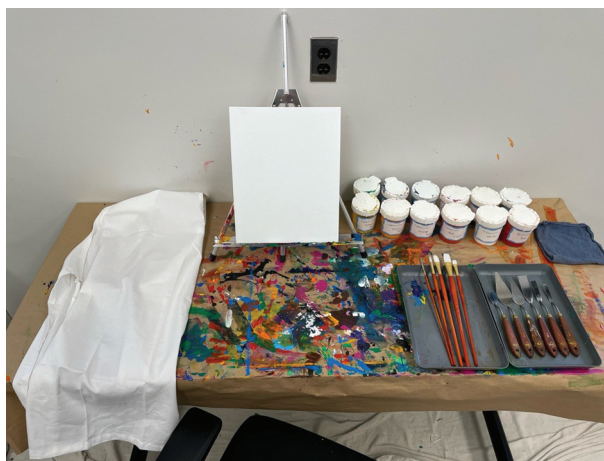

Figure 1. Set-up of painting task, which included paintbrushes, paints, palette knives, a canvas, and an apron.

were then given 15 minutes to paint as they wished within the time, the only instruction being to produce an abstract painting. Abstract painting was defined to participants as such: " … a painting that does not represent images of our everyday world. It has lines and shapes, but they are not meant to represent objects or living things."

Ratings were done by five human raters (all of whom were undergraduate students interested in creativity research), who were individually shown photographs of all 99 paintings. They were then asked to simply rate each painting, on a scale of 1 (not at all creative) through 7 (extremely creative); each participant thus rated each painting once. To determine initial interrater reliability metrics before implementing MFRM, we performed Light's (1971) using Fleiss-Cohen's weighted Kappa (Fleiss, Cohen, & Everitt, 1969). For these Painting Creativity scores, $\kappa = .37$, which is considered fair agreement per Landis and Koch (1977) and on par with previous work examining subjective ratings in creativity scores (Primi, Silvia, Jauk, & Benedek, 2019). Given the satisfactory interrater reliability and lack of any major rating difficulties, we implemented the MFRM, in which we specified participants and raters as facets. We estimated Rasch "fair average" scores, which were used as our overall measure of painting ability. Our five raters varied in their severity, from −0.17 to 0.67 (less to more severe; in the Z metric), thus justifying the use of faceted Rasch scaling. The scores are scores after adjusting for the difficulty of the severity of the raters ranged from −1.90 to 1.68 (in the Z metric). The model's EAP reliability was 0.78, indicating that the MFRM scores reliably differentiated people's underlying creative painting ability; we reach this conclusion as EAP quantifies the internal consistency of the scores accounting for the design (both items and raters) rather than quantifying rater consistency across people. The many-facet Rasch analyses were carried out using *TAM* package (version 3.7–16; Robitzsch, Kiefer, & Wu, 2021) in R (version 4.1.0; R Core Team, 2021).

Test of Creative Imagery Abilities (forms A and B).
Participants all received one of two paper versions of the TCIA (Jankowska & Karwowski, 2015), a figural creativity task that asked participants to draw pictures using 7 image prompts. The prompts were combinations of 2–4 elements, either dots or lines (see Appendix A for sample images). Form B was simply a rotation of Form A by 180 degrees. Participants were instructed to underline the idea they liked the most and to draw and title their idea. Then, they were told that they could complete the image with unrestricted elements, as well as change and develop it to create something even more unusual.

Three undergraduate raters were trained on the scoring scheme, which targeted three qualities of the finished product: Vividness, Originality, and Transformativeness. A high level of Vividness was recognized by an abundance of detail in the completion of the initial figure, a clear depiction of motion and dynamics in the drawing, or a complex presentation of metaphorical and symbolic content. A high level of Originality was recognized by a depiction of new objects, activities, processes, and events in the drawing that differ considerably from the actually existing ones, surprising and novel presentation of cultural artifacts such as works of art, or amusing presentation of contents, suggesting a good sense of humor. A high level of Transformativeness required multiplication (multiplying an element of the image), hyperbolization (excessive distortion of proportions, for example by emphasizing an element of the image), or amplification (adding detail to the image). Participants received 0 to 2 points in each category for each drawing.

Similar to the abstract painting task, we estimated Fleiss-Cohen's weighted Kappa as a metric of interrater reliability for each TCIA rating (Vividness, Originality, and Transformativeness). Agreement ranged from fair to moderate (Landis & Koch, 1977): For Vividness, $\kappa = 0.42$; for Originality, $\kappa = 0.33$; and for Transformativeness, $\kappa = .22$. We then estimated a MFRM to adjust for rater severity but also prompt severity for each TCIA rating. We specified participants, raters, and prompts as facets. Our three raters varied in their severity, from −0.65 to 0.49 (less severe to more severe; in the Z metric) for the Vividness rating, −0.73 to 0.89 for the Originality rating, and −0.87 to 0.55 for the Transformativeness rating. These varying severities justify the use of faceted Rasch scaling. The scores are generated after adjusting for the difficulty of the prompts and severity of the raters, which ranged from −2.50 to 2.33 (in the Z metric) for the Vividness rating, −2.29 to 2.46 for the Originality rating, and −1.93 to 2.27 for the Transformativeness rating. The model's Rasch reliability was 0.89 for Vividness, 0.84 for Originality, and 0.82 for Transformativeness, indicating that MFRM scores reliably differentiated people's underlying abilities.

Alternate Uses Task.
To test verbal divergent-thinking ability, participants were given three rounds of a computerized AUT (Guilford, 1967), presented on Qualtrics survey software. They were asked to think of as many original and creative uses for prompted objects as they could, and were encouraged to come up with responses that "strike people as clever, unusual, interesting, uncommon, humorous, innovative, or different." They responded to the three prompts ("box," "rope," and "pen") for 2 minutes each.

We instructed three undergraduate raters to use the subjective scoring method in assigning creative scores for AUT responses. As such, they were guided about what a "creative response" could reflect (e.g., original, novel, interesting), but these terms relied on the rater's subjective concept of such terms (Beaty & Johnson, 2021; Cseh & Jeffries, 2019). The raters assigned each response a value from 1 ("not at all creative") to 5 ("very creative"), using similar language to how paintings were rated for creativity. The raters then reconciled the ratings that diverged the most. This scoring method has shown to be valid and reliable in past studies (Jauk, Benedek, & Neubauer, 2014; Silvia et al., 2008).

Similar to the TCIA, we first estimated interrater reliability with Fleiss-Cohen's weighted Kappa; for AUT scores, $\kappa = .35$ (fair agreement; Landis & Koch, 1977), in line with some past work assessing subjective AUT ratings (Primi, Silvia, Jauk, & Benedek, 2019). We then estimated a MFRM to adjust for rater severity and prompt severity. In this model, each person's estimated trait divergent-thinking ability is adjusted for variation in the difficulty of the prompts and differences in the raters' severity. We specified participants, raters, and prompts as facets. We estimated Rasch "fair average" scores, which were used as our overall measure of creative painting ability. Our three raters varied in their severity, from −0.90 to 0.87 (less severe to more severe; in the Z metric), thus justifying the use of faceted Rasch scaling. The scores are scores after adjusting for the difficulty of the prompts and severity of the raters and ranged from −1.08 to 1.21 (in the Z metric). The model's Rasch reliability was 0.88, indicating the MFRM scores reliably differentiated people's underlying divergent thinking ability.

Fluid Intelligence Scale.
To assess fluid intelligence (Gf), participants were given 3 minutes to get through as many sequence-completion questions as they could (Cattell & Cattell, 1973). For each of the 13 questions, three pictures are shown, and participants must select a fourth image, out of 6 options, to complete the sequence. Scores were aggregated through summation.

Exploratory measures
Forward flow task.
Forward flow (FF) is a measure that is adjacent to verbal divergent-thinking ability (Gray et al., 2019). It aims to measure participants' ability to connect concepts that, theoretically, are disparately related to one another in their semantic space. Our administration of the task required participants to fill 19 text boxes with words that were serially, semantically connected to the preceding word. The task began with a prompt word, of which there were three ("bear," "candle," and "table"). Final FF scores were averaged across these three trials.

Inventory of Creative Activities and Achievements.
Another measure of creative engagement (ICAA) was collected by presenting participants with 65 activities across eight domains of creativity: literature, music, arts and crafts, creative cooking, sports, visual arts, performing arts, and science and engineering (Diedrich et al., 2018). For each domain, participants reported on a scale from "never" to "more than 10 times" to indicate how often they had carried out listed activities over the last 10 years. Participants also provided the number of years they estimated to have spent engaged in each domain. Additionally, the participants were asked to list their five most creative achievements in their lives. Overall "Creative Activities" and "Creative Achievements" scores and domain-specific scores were aggregated through summation.

Creative Mindset Scale.
The Creative Mindset Scale (CMS) measured whether participants' beliefs about their own creative abilities reflected a "growth mindset" or a "fixed mindset" (Karwowski, 2014). An example of a belief that creativity is a primarily innate ability ("fixed mindset") is, "Creativity can be developed, but one either is or is not a truly creative person," whereas an example of a belief that creativity can be developed ("growth mindset") is, "Practice makes perfect – perseverance and trying hard are the best ways to develop and expand one's capabilities." These items were rated on a scale of 1 ("definitely not") to 5 ("definitely yes"). Scores for each mindset were aggregated through summation.

NEO-Five Factor Personality Inventory.
The 60-item NEO-FFI (Costa & McCrae, 1992) was selected to assess the five major personality factors: Openness to Experience, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. Participants indicated their agreement with the items on a 5-point scale from 1 ("strongly disagree") to 5 ("strongly agree"). Scores for each factor were aggregated through summation, reverse-scored when applicable. See Appendix B for hierarchical omega scores for each factor.

Procedure
Upon arrival, participants were called into the testing room and directed to a computer station, where they provided consent before continuing. The participants then completed the six non-drawing measures on the computer via Qualtrics, with presentation order randomized. After completion, participants alerted the experimenter, who administered the paper version of the

Table 1. Pearson's correlations.

	Painting Creativity	Vividness	Originality	Transform.	AUT	Gf
1. Painting Creativity	3.87 (0.96)					
2. Vividness	.175	0.85 (0.53)				
3. Originality	.281**	.702***	1.12 (0.50)			
4. Transformativeness	.202*	.751***	.807***	0.80 (0.47)		
5. AUT	.077	.232*	.158	.165	1.87 (0.36)	
6. Gf	.053	.322**	.338**	.223*	.293**	8.56 (1.60)

* $p < .05$, ** $p < .005$, *** $p < .001$; Means and standard deviations (in parentheses) are on the diagonal.

TCIA. Participants were then given the final 15 minutes to work freely on an abstract painting. They were then thanked and compensated for their time.

Results

Primary creativity measures of interest

Painting Creativity scores were positively correlated with Originality ($r = .281$, $p = .005$) and Transformativeness ($r = .202$, $p = .045$) scores on the TCIA, but did not correlate with performance on the AUT ($r = .077$, $p = .449$; see Table 1).[1] All three sub-scores of the TCIA were strongly intercorrelated. The AUT and TCIA only correlated with the Vividness score of the TCIA ($r = .232$, $p = .021$). Fluid Intelligence was significantly positively correlated with the AUT ($r = .293$, $p = .003$), Vividness ($r = .322$, $p = .001$), Originality ($r = .338$, $p < .001$), and Transformativeness ($r = .223$, $p = .027$).

Exploratory measures

A full correlation matrix can be found in Appendix C. The correlation between Painting Creativity and Conscientiousness was marginally negatively significant, at $r = -.195$ ($p = .05$). Painting Creativity did not correlate significantly with any other exploratory measure.

All three subscores of the TCIA correlated positively with Creative Activities (Originality: $r = .307$, $p = .002$; Vividness: $r = .260$, $p = .009$; Transformativeness: $r = .217$, $p = .031$) and Achievements (Originality: $r = .354$, $p < .001$; Vividness: $r = .312$, $p = .002$; Transformativeness: $r = .242$, $p = .016$).

Extraversion was significantly correlated with engagement with Creative Activities ($r = .324$, $p = .001$), but not Achievements. Neuroticism was negatively correlated with Growth mindset ($r = -.208$, $p < .05$). Agreeableness and Openness were both positively correlated with Growth mindset (Agreeableness: $r = .315$, $p = .001$; Openness: $r = .257$, $p = .01$) and strongly, negatively correlated with Fixed mindset

(Agreeableness: $r = -.414$, $p < .001$; Openness: $r = -.297$, $p = .003$). Openness also correlated with Creative Activities ($r = .451$, $p < .001$) and Achievements ($r = .318$, $p = .001$). Forward Flow did not associate with any variable.

Openness to Experience item-level correlations

We sought to specifically investigate the Openness dimension of personality at the item-level. In contemporary personality literature, item-level data can have more predictive power than traits (e.g., Seeboth & Mõttus, 2018). This is likely because, when calculating a trait, true relationships evident in item-level data may become obscured through collapsing of the items. In our item-level analysis of Openness, the item most widely associated with creative performance was question O3: "I am intrigued by the patterns I find in art and nature," which was positively correlated with Painting Creativity ($r = .254$, $p = .011$), TCIA Vividness ($r = .221$, $p = .028$), and TCIA Originality ($r = .280$, $p = .005$). Question O2, "I think it's interesting to learn and develop new hobbies," was positively correlated with TCIA Vividness ($r = .306$, $p = .002$), TCIA Originality ($r = .224$, $p = .026$), and TCIA Transformativeness ($r = .254$, $p = .011$). Reverse-scored O7, "I seldom notice the moods or feelings that different environments produce," positively correlated with Painting Creativity ($r = .235$, $p = .019$). O9, "Sometimes when I am reading poetry or looking at a work of art, I feel a chill or wave of excitement," correlated positively with TCIA Originality ($r = .230$, $p = .022$).

Discussion

This study demonstrates that having people produce freeform abstract paintings provides an alternate index of visual creativity that merits future research. So as not to constrain participants' painting process too much, our instructions were minimal: produce a painting in fifteen minutes that is abstract, i.e., it does not represent images of our everyday world. All participants except one followed our instructions

and produced unique works of art. When five experimenters independently rated these works, two important findings emerged. First, raters, in the absence of being given specific criteria, were reliable in judging painting quality ("how creative do you find this piece?"), both before implementation of MFRM and with the MFRM. Second, the average creativity ratings given to these paintings were normally distributed (see Appendix D; skewness = .09, kurtosis = −.31). Taken together, our results indicate that this task captures creative differences in abstract paintings that vary meaningfully across the general population.

We were interested in whether two "gold standard" creativity tasks – the AUT and the TCIA – captured a domain-general ability that generalized to the painting task, or if performance on all three tasks was differentiated and domain-specific. We found some evidence that performance on the TCIA, a drawing task, more closely tracked with painting performance than the AUT, a written task, which did not correlate. However, AUT score did correlate with the Vividness sub-score of the TCIA ($r = .232$). If this result is replicated, it might indicate that the TCIA and AUT both make use of individuals' visual imagery capacities, but in different ways. However, although the subjective scoring method has been shown to be valid and reliable (Jauk, Benedek, & Neubauer, 2014; Silvia et al., 2008), potential inherent subjectivity and rater exhaustion could affect ratings, which may influence overall AUT associations with other gold standard tasks (Forthmann et al., 2017).

Our exploratory findings should be interpreted cautiously since the study was not designed to test a priori hypotheses. However, our personality, mindset, and performance findings either replicate previous work or add to a small body of knowledge that investigated these correspondences. For example, Openness correlated with Creative Activities ($r = .451$), Creative Achievements ($r = .318$), and a Growth mindset of creativity ($r = .257$). While Openness would be expected to correlate positively with Originality on the TCIA (or figural creativity, more generally) based on predictions by McCrae (1987), to our knowledge, no prior research examined the intersection of personality and figural creativity directly.

Another exploratory finding related to personality and creativity was the negative relationship between Conscientiousness and Painting Creativity ($r = −.195$). The relationship between conscientiousness and creativity was examined in a meta-analysis by Reiter-Palmon, Illies, and Kobe-Cross (2009), who found that splitting Conscientiousness into "Achievement" and "Dependability" yielded a positive relationship with

creative performance with the former, and a negative relationship with the latter. However, there is little research in this area.

One of the more theoretically laden findings among the exploratory variables was the positive correlation between Fluid Intelligence and both laboratory measures of creativity performance. The finding that Fluid Intelligence correlated strongly with the AUT and all three sub-scores of the TCIA, but not with Painting Creativity, adds to the mounting evidence that the former two creativity assessments target a capacity that resembles domain-general intelligence (Frith et al., 2021; Silvia, 2015) or may reflect a combination of shared capacities for attentional control and verbal intelligence (Benedek et al., 2017; Frith et al., 2021). Conversely, another interpretation is that, given that the painting task correlates with the TCIA but not with other performance measures, this painting task yields a more "distilled" measure of domain-specific creativity.

Despite each measure bearing at least one relationship consistent with previous reports, there was also a notable lack of correspondence between some of these constructs. Despite correlating positively with TCIA Originality, Creative Activities, and Creative Achievements, Openness did not correlate with performance on the AUT in this study. This finding is surprising, as this relationship is found consistently enough that it is used to infer accuracy of computerized scoring techniques relative to humans (Acar et al., 2021; Beaty & Johnson, 2021). Sampling variability and rater fatigue are potential sources for this difference but require further research (Forthmann et al., 2017). In addition, Forward Flow did not associate significantly with any variables, even though it has been increasingly used in models of creative cognition (Beaty, Zeitlen, Baker, & Kenett, 2021); this lack of association merits future research. Furthermore, Painting Creativity did not correspond to Creative Activities or Achievements, even when the Visual Arts practice subscore was isolated and compared (see Appendix C). The absence of this relationship should not be interpreted as no relationship but instead a null finding that warrants additional research. However, it is possible that the quality of amateur paintings is unrelated to the amount of time individuals spend practicing their visual arts skills.

A distinct, but related, consideration involves the use of non-expert painting raters. Although we did not set out to use the Consensual Assessment Technique to score these paintings, our protocol resembles a non-expert execution of the protocol (though with alternate resources in a novel artistic domain, affording a different reflection of a participant's creativity output). We

felt confident in the ability of non-expert raters based on past work that has found high correlations between novice versus expert raters across creative domains (e. g., in short story ratings, *r* = .71, Kaufman, Baer, & Cole, 2009; in judgments of children's drawings, *r* = .69; Amabile, 1982). In addition, Hawley-Dolan and Winner (2011) found similarities between experts and non-expert rating patterns of abstract paintings specifically. To support this point, we also collected ratings from an expert painter, which found very similar results in conjunction with our non-expert ratings, with similarities in rating tendencies as well. That is, an expert rater maintained very similar range and severity of scores, and when aggregated with non-expert ratings, increased reliability and replicated non-significant correlations with the AUT gold-standard. Such findings corroborate the past work where non-experts' and experts' creativity ratings have correlated highly, despite Kaufman, Baer, Cole, and Sexton (2008) caution against the free replacement of experts with non-experts based on sometimes low associations between the two groups (e.g., poetry assessment in Kaufman, Baer, Cole, & Sexton, 2008; flower design judgments in Lee, Lee, & Youn, 2005). However, in related tasks like aesthetic judgments, existing work has even shown that those with less artistic experience produce more consistent ratings of some attributes of museum-grade paintings than those with moderate levels of experience, though in general no significant differences exist between the two (Chatterjee et al., 2010). Similarly, Hawley-Dolan and Winner (2011) found that non-experts preferred and judged professional artworks strikingly similar to experts, even with misleading labels. Whether rating the creativity of a painting is more subjective than specific aspects of a painting is an open question beyond the scope of our primary examinations into abstract paintings as a creative measure, but our evidence suggests that non-experts can reach considerable consensus.

Ultimately, future work is merited to determine the true relatedness of the two groups, particularly in painting. One future direction could implement Kaufman and Baer's (2012) distinction between experts, non-experts, and "quasi-experts" (though the definition of the term varies by study). Seen as a middle-ground between the knowledge of experts but practical ease of using non-experts, exactly who is considered a "quasi-expert" has varied widely, from IMDb users judging films (Plucker et al., 2008), creativity psychologists judging short literature (Baer, Kaufman, & Gentile, 2004), or teachers with a decade of instruction practice assessing poems (Cheng, Wang, Liu, & Chen, 2010). Thus, it is possible that our undergraduate raters with research experience in creativity psychology fall into this category. Regardless, in this study, given that our raters achieved acceptable reliability for rating the paintings collected here, that we implemented MFRMs to adjust for rater severity, and that the ratings of an expert mirrored those of the non-experts, our use of non-expert raters is not necessarily cause for concern.

There are several ways to improve this protocol in future work. First, providing "be creative" instructions has been shown to increase people's divergent-thinking ability on the AUT and may similarly improve people's creativity in their abstract paintings, if such language were added to future painting instructions (Nusbaum, Silvia, & Beaty, 2014). Second, intuitively and empirically, taking longer to complete a creativity task may improve performance. Acar et al. (2021) found in their meta-analysis of 1325 verbal and 488 figural responses that longer think time predicted originality, across different divergent-thinking tasks. Future studies could allot 3–5 minutes to complete the AUT and encourage participants to take the full 20 minutes to complete the seven images of the TCIA. We could also extend the time allotted to the painting task.

Given that the general "creativity ratings" were internally consistent and varied meaningfully, an interesting avenue for future work would be to explore different rating criteria for abstract paintings and identify separate factors or subscores, as in the TCIA. This would also help elucidate the aforementioned question on the degree of subjectivity regarding creativity versus other properties of a painting, like beauty. One possibility is to use the same scoring criteria as the TCIA. However, given that freely painted abstract art may carry more emotional associations than simple figural designs, a starting point for developing additional rating questions might consider the list of 11 "fundamental terms" to describe reactions to art that were distilled by Anjan Chatterjee and colleagues (Chatterjee, 2020; Christensen, Cardillo, & Chatterjee, 2023). Future work could determine whether any of these words could be used to assess creative aptitude or go through the process of creating a similar list of indicators of creativity with which to rate paintings. Another possibility is to use the 11 fundamental terms, many of which describe affective reactions to art, to train raters on how to judge the "expressiveness" of a piece.

Relatedly, experimenters could directly target the affective experiences of abstract painting and determine its potential role in moderating resultant creative quality. Along these lines, Kharkhurin (2014) briefly discusses the potential for abstract painting to elicit mood effects without the skill of representational painting. Although speculative, we think that commonly used

tasks such as the AUT would have little, if any, benefit to one's well-being, particularly when compared to more face-valid creativity tasks such as abstract painting. Thus, this paradigm that might reasonably permit us to, in future research, explore the potential benefits of applied creative tasks like abstract painting on well-being.

Beyond further refinement of the protocol itself, and increasing sophistication of the rating schemes, there is significant room for further exploration into how creative performance on abstract freeform painting predicts performance in other creative domains and in real-world creative achievement. For instance, we did not implement the common Torrance Test of Creative Thinking. With its "Figural" component, in future work we could perhaps clarify how abstract painting associates with other creative tasks like this. In addition, in the current literature that uses the AUT as a proxy for creative potential, the relationships between creative ability and creative achievement are not straightforward. For example, creative achievement may require an interplay between the Openness personality trait, forms of intelligence, motivation, and domain-specific expertise (Jauk, Benedek, & Neubauer, 2014). As such, we cannot dismiss the possibility that paintings reflect creative potential perhaps in more complex interactions with individual traits or in specific creative domains.

In looking back to Guilford's (1950) call to find means by which to measure creativity as an ability independent of intelligence, our abstract, freeform painting task seems to accomplish that aim, and yield reliable, normally distributed scores of creativity. Painting represents a novel, yet familiar, means to apprehend the nature of creative capacity more deeply. Surely, such a task is not as easy to implement or score as traditional laboratory-based tasks that have dominated the literature. However, based on our results, we think this protocol represents a fruitful avenue for future work in collecting and rating a sample of ecologically and face-valid creative products. Inclusion of a painting task helps illuminate the similarities and differences between performance measures, including text-based and figural divergent-thinking tasks, fluid intelligence tests, and hands-on artistic activities, and offers us another path to advancing understanding.

Note

1. We also explored how ratings of an expert (i.e., a professional painter) on Painting Creativity scores

would affect results. With expert ratings, κ increases from .37 to .38, and EAP reliability increases from 0.78 to 0.81, indicating consistency between the two groups of raters. Before and after addition of expert ratings, rater severity remained similar ([−0.17, 0.67] to [−0.36, 0.61]) and range of scores remained similar ([−1.90, 1.68] to [−2.09, 1.60]). Importantly, correlation with the AUT remains non-significant, from $r = .077$ to $r = .093$.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Lucas Bellaiche ⓘD http://orcid.org/0000-0002-5548-4271
Anna P. Smith ⓘD http://orcid.org/0000-0002-1365-0809

Declaration statement

The authors report there are no competing interests to declare.

References

Acar, S., Berthiaume, K., Grajzel, K., Dumas, D., Flemister, C. T., & Organisciak, P. (2021). Applying automated originality scoring to the verbal form of Torrance tests of creative thinking. *The Gifted Child Quarterly*, *67*(1), 3–17. doi:10.1177/00169862211061874

Alabbasi, A. M. A., Paek, S. H., Kim, D., & Cramond, B. (2022). What do educators need to know about the Torrance tests of creative thinking: A comprehensive review. *Frontiers in Psychology*, *13*, 1000385. doi:10.3389/fpsyg.2022.1000385

Amabile, T. M. (1982). Social psychology of creativity: A consensual assessment technique. *Journal of Personality and Social Psychology*, *43*(5), 997. doi:10.1037/0022-3514.43.5.997

Ausubel, D. P. (1967). *Learning theory and classroom practice*. Toronto: Ontario Institute for Studies in Education Bulletin.

Baer, J., Kaufman, J. C., & Gentile, C. A. (2004). Extension of the consensual assessment technique to nonparallel creative products. *Creativity Research Journal*, *16*(1), 113–117. doi:10.1207/s15326934crj1601_11

Barron, F. (1965). Some studies of creativity at the institute of personality assessment and research. In G. Steiner (Ed.), *The creative organization* (pp. 118–125). Chicago, IL: University of Chicago Press.

Beaty, R. E. (2015). The neuroscience of musical improvisation. *Neuroscience & Biobehavioral Reviews*, *51*, 108–117. doi:10.1016/j.neubiorev.2015.01.004

Beaty, R. E., & Johnson, D. R. (2021). Automating creativity assessment with SemDis: An open platform for computing semantic distance. *Behavior Research Methods*, *53*(2), 757–780. doi:10.3758/s13428-020-01453-w

Beaty, R. E., & Silvia, P. J. (2013). Metaphorically speaking: Cognitive abilities and the production of figurative language. *Memory & Cognition*, 41(2), 255–267. doi:10.3758/s13421-012-0258-5

Beaty, R. E., Zeitlen, D. C., Baker, B. S., & Kenett, Y. N. (2021). Forward flow and creative thought: Assessing associative cognition and its role in divergent thinking. *Thinking Skills and Creativity*, 41, 100859. doi:10.1016/j.tsc.2021.100859

Benedek, M., Kenett, Y. N., Umdasch, K., Anaki, D., Faust, M., & Neubauer, A. C. (2017). How semantic memory structure and intelligence contribute to creative thought: A network science approach. *Thinking & Reasoning*, 23(2), 158–183. doi:10.1080/13546783.2016.1278034

Carson, S. H., Peterson, J. B., & Higgins, D. M. (2005). Reliability, validity, and factor structure of the creative achievement questionnaire. *Creativity Research Journal*, 17(1), 37–50. doi:10.1207/s15326934crj1701_4

Cattell, R. B., & Cattell, A. K. (1973). *Measuring intelligence with the culture fair tests*. Champaign, IL: Institute for Personality and Ability Testing.

Chatterjee, A. (2020, July 1). *Coming to Terms with Art*. Templeton Religion Trust. https://templetonreligiontrust.org/explore/coming-to-terms-with-art/

Chatterjee, A., Widick, P., Sternschein, R., Smith, W. B., & Bromberger, B. (2010). The assessment of art attributes. *Empirical Studies of the Arts*, 28(2), 207–222. doi:10.2190/EM.28.2.f

Cheng, Y.-Y., Wang, W.-C., Liu, K.-S., & Chen, Y.-L. (2010). Effects of association instruction on fourth graders' poetic creativity in Taiwan. *Creativity Research Journal*, 22(2), 228–235. doi:10.1080/10400419.2010.481542

Christensen, A. P., Cardillo, E. R., & Chatterjee, A. (2023). Can art promote understanding? A review of the psychology and neuroscience of aesthetic cognitivism. *Psychology of Aesthetics, Creativity, and the Arts*. doi:10.1037/aca0000541

Christensen, P. R., & Guilford, J. P. (1963). An experimental study of verbal fluency factors. *British Journal of Statistical Psychology*, 16(1), 1–26. doi:10.1111/j.2044-8317.1963.tb00195.x

Christensen, A. P., Silvia, P. J., Nusbaum, E. C., & Beaty, R. E. (2018). Clever people: Intelligence and humor production ability. *Psychology of Aesthetics, Creativity, and the Arts*, 12(2), 136–143. doi:10.1037/aca0000109

Costa, P. T., Jr., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual*. Odessa, FL: Psychological Assessment Resources.

Cseh, G. M., & Jeffries, K. K. (2019). A scattered CAT: A critical evaluation of the consensual assessment technique for creativity research. *Psychology of Aesthetics, Creativity, and the Arts*, 13(2), 159–166. doi:10.1037/aca0000220

Diedrich, J., Jauk, E., Silvia, P. J., Gredlein, J. M., Neubauer, A. C., & Benedek, M. (2018). Assessment of real-life creativity: The Inventory of Creative Activities and Achievements (ICAA). *Psychology of Aesthetics, Creativity, and the Arts*, 12(3), 304–316. doi:10.1037/aca0000137

Eckes, T. (2011). Introduction to many-facet rasch measurement. In *Frankfurt am. Main*: *Peter Lang*. 10.3726/978-3-653-04844-5.

Fleiss, J. L., Cohen, J., & Everitt, B. S. (1969). Large sample standard errors of kappa and weighted kappa. *Psychological Bulletin*, 72(5), 323–327. doi:10.1037/h0028106

Forthmann, B., Holling, H., Zandi, N., Gerwig, A., Çelik, P., Storme, M., & Lubart, T. (2017). Missing creativity: The effect of cognitive workload on rater (dis-)agreement in subjective divergent-thinking scores. *Thinking Skills and Creativity*, 23, 129–139. doi:10.1016/j.tsc.2016.12.005

Frith, E., Elbich, D. B., Christensen, A. P., Rosenberg, M. D., Chen, Q., Kane, M. J., & Beaty, R. E. (2021). Intelligence and creativity share a common cognitive and neural basis. *Journal of Experimental Psychology: General*, 150(4), 609–632. doi:10.1037/xge0000958

Frith, E., Kane, M. J., Welhaf, M. S., Christensen, A. P., Silvia, P. J., & Beaty, R. E. (2021). Keeping creativity under control: Contributions of attention control and fluid intelligence to divergent thinking. *Creativity Research Journal*, 33(2), 138–157. doi:10.1080/10400419.2020.1855906

Gray, K., Anderson, S., Chen, E. E., Kelly, J. M., Christian, M. S. … Lewis, K. (2019). "Forward flow": A new measure to quantify free thought and predict creativity. *The American Psychologist*, 74(5), 539–554. doi:10.1037/amp0000391

Guilford, J. P. (1950). Creativity. *American Psychologist*, 5(9), 444–454. doi:10.1037/h0063487

Guilford, J. P. (1967). *The nature of human intelligence*. New York, NY: McGraw-Hill.

Halasz, P. (2009). *A memoir of creativity: Abstract painting, politics & the media, 1956-2008*. Bloomington, IN: iUniverse.

Hawley-Dolan, A., & Winner, E. (2011). Seeing the mind behind the art: People can distinguish abstract expressionist paintings from highly similar paintings by children, chimps, monkeys, and elephants. *Psychological Science*, 22(4), 435–441. doi:10.1177/0956797611400915

Hennessey, B. A., Amabile, T. M., & Mueller, J. S. (2011). In M. A. Runco & S. R. Pritzker (Eds), *Encyclopedia of creativity* (pp. 253–260). San Diego, CA: Academic Press. doi:10.1016/B978-0-12-375038-9.00046-7

Jankowska, D. M., & Karwowski, M. (2015). Measuring creative imagery abilities. *Frontiers in Psychology*, 6, 1591. doi:10.3389/fpsyg.2015.01591

Jauk, E., Benedek, M., Dunst, B., & Neubauer, A. C. (2013). The relationship between intelligence and creativity: New support for the threshold hypothesis by means of empirical breakpoint detection. *Intelligence*, 41(4), 212–221. doi:10.1016/j.intell.2013.03.003

Jauk, E., Benedek, M., & Neubauer, A. C. (2014). The road to creative achievement: A latent variable model of ability and personality predictors. *European Journal of Personality*, 28(1), 95–105. doi:10.1002/per.1941

Jellen, H. G., & Urban, K. K. (1989). Assessing creative potential world-wide: The first cross-cultural application of the test for creative thinking—drawing production (TCT-DP). *Gifted Education International*, 6(2), 78–86. doi:10.1177/026142948900600204

Johnson, D. R., Kaufman, J. C., Baker, B. S., Patterson, J. D., Barbot, B. … Beaty, R. E. (2022). Divergent semantic integration (DSI): Extracting creativity from narratives with distributional semantic modeling. *Behavior Research Methods*. doi:10.3758/s13428-022-01986-2

Karwowski, M. (2014). Creative mindsets: Measurement, correlates, consequences. *Psychology of Aesthetics, Creativity, and the Arts*, 8(1), 62–70. doi:10.1037/a0034898

Karwowski, M., Dul, J., Gralewski, J., Jauk, E., Jankowska, D. M. … Benedek, M. (2016). Is creativity without intelligence possible? A necessary condition analysis. *Intelligence*, 57, 105–117. doi:10.1016/j.intell.2016.04.006

Kaufman, J., & Baer, J. (2012). Beyond new and appropriate: Who decides what is creative? *Creativity Research Journal*, 24(1), 83–91. doi:10.1080/10400419.2012.649237

Kaufman, J. C., Baer, J., & Cole, J. C. (2009). Expertise, domains, and the consensual assessment technique. *The Journal of Creative Behavior*, 43(4), 223–233. doi:10.1002/j.2162-6057.2009.tb01316.x

Kaufman, J. C., Baer, J., Cole, J. C., & Sexton, J. D. (2008). A comparison of expert and nonexpert raters using the consensual assessment technique. *Creativity Research Journal*, 20(2), 171–178. doi:10.1080/10400410802059929

Kharkhurin, A. (2014). Creativity.4in1: Four-criterion construct of creativity. *Creativity Research Journal*, 26(3), 338–352. doi:10.1080/10400419.2014.929424

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174. doi:10.2307/2529310

Lee, S., Lee, J., & Youn, C.-Y. (2005). A variation of CAT for measuring creativity in business products. *Korean Journal of Thinking & Problem Solving*, 15, 143–153.

Light, R. J. (1971). Measures of response agreement for qualitative data: Some generalizations and alternatives. *Psychological Bulletin*, 76(5), 365–377. doi:10.1037/h0031643

Linacre, J. M. (1994). *Many-faceted Rasch measurement* (2nd ed.). Chicago, IL: University of Chicago.

McCrae, R. R. (1987). Creativity, divergent thinking, and openness to experience. *Journal of Personality and Social Psychology*, 52(6), 1258–1265. doi:10.1037/0022-3514.52.6.1258

Nusbaum, E. C., Silvia, P. J., & Beaty, R. E. (2014). Ready, set, create: What instructing people to "be creative" reveals about the meaning and mechanisms of divergent thinking. *Psychology of Aesthetics, Creativity, and the Arts*, 8(4), 423–432. doi:10.1037/a0036549

Nusbaum, E. C., Silvia, P. J., & Beaty, R. E. (2017). Ha ha? Assessing individual differences in humor production ability. *Psychology of Aesthetics, Creativity, and the Arts*, 11(2), 231–241. doi:10.1037/aca0000086

Plucker, J. A., Holden, J., & Neustadter, D. (2008). The criterion problem and creativity in film: Psychometric characteristics of various measures. *Psychology of Aesthetics, Creativity, and the Arts*, 2(4), 190–196. doi:10.1037/a0012839

Primi, R., Silvia, P. J., Jauk, E., & Benedek, M. (2019). Applying many-facet rasch modeling in the assessment of creativity. *Psychology of Aesthetics, Creativity, and the Arts*, 13(2), 176–186. doi:10.1037/aca0000230

R Core Team. (2021). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/

Reiter-Palmon, R., Illies, J. J., & Kobe-Cross, L. M. (2009). Conscientiousness is not always a good predictor of performance: The case of creativity. *The International Journal of Creativity & Problem Solving*, 19(2), 27–45.

Robitzsch, A., Kiefer, T., & Wu, M. (2021). *TAM: Test Analysis Modules. R package version 3.7-16*. Retrieved from https://CRAN.R-project.org/package=TAM

Runco, M. A., & Acar, S. (2012). Divergent thinking as an indicator of creative potential. *Creativity Research Journal*, 24(1), 66–75. doi:10.1080/10400419.2012.652929

Seeboth, A., & Mõttus, R. (2018). Successful explanations start with accurate descriptions: Questionnaire items as personality markers for more accurate predictions. *European Journal of Personality*, 32(3), 186–201. doi:10.1002/per.2147

Silvia, P. J. (2015). Intelligence and creativity are pretty similar after all. *Educational Psychology Review*, 27(4), 599–606. doi:10.1007/s10648-015-9299-1

Silvia, P. J., & Beaty, R. E. (2012). Making creative metaphors: The importance of fluid intelligence for creative thought. *Intelligence*, 40(4), 343–351. doi:10.1016/j.intell.2012.02.005

Silvia, P. J., Christensen, A. P., & Cotter, K. N. (2021). Right-wing authoritarians aren't very funny: RWA, personality, and creative humor production. *Personality and Individual Differences*, 170, 110421. doi:10.1016/j.paid.2020.110421

Silvia, P. J., Winterstein, B. P., Willse, J. T., Barona, C. M., Cram, J. T. … Richard, C. A. (2008). Assessing creativity with divergent thinking tasks: Exploring the reliability and validity of new subjective scoring methods. *Psychology of Aesthetics, Creativity, and the Arts*, 2(2), 68–85. doi:10.1037/1931-3896.2.2.68

Torrance, P. (1966). *Torrance tests of creative thinking*. Princeton, NJ: Personnel Press.

Appendices

Appendix A: Measures Used

Alternate Uses Task

Instructions:

For this task, you'll be asked to come up with as many original and creative uses for objects as you can. The goal is to come up with *creative ideas*, which are ideas that strike people as clever, unusual, interesting, uncommon, humorous, innovative, or different.

You will be asked to type uses for 3 different objects.

You will have 2 minutes to type as many creative uses for each object as you can – just press TAB after each one.

Click the arrow to begin.

Example Prompt: Please list all of the creative, unusual uses for a ROPE you can think of.

Press TAB after each one.

Forward Flow Task

Instructions:

In this task, starting from a given word, your job is to write down the next word that follows in your mind from the previous word. Please put down only single words, and do not use proper nouns (such as names, brands, etc.).

Press TAB after each one.

Click the arrow to begin.

Example Prompt: Write down the next word that follows in your mind from the previous word.

Press TAB after each word. Continue when all text boxes are complete.

Your starting word is ["Bear," "Table," "Candle"]

Test of Creative Image Abilities

Each participant receives 7 image prompts, either A or B, each with the following instructions:
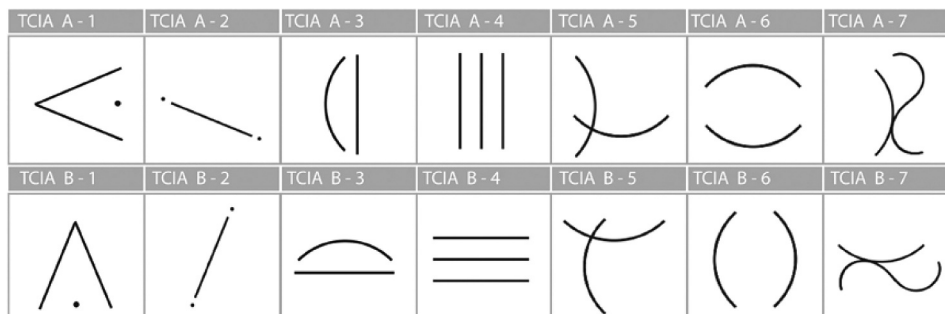
What does this drawing remind you of? Please, write down.

The more ideas, the better!

Inside the box, underline the idea that you like the most. You can complete it with unrestricted elements, change and develop it so you create something even more unusual. Please, draw your idea.

Write down the title.

Image Prompts:

Fluid Intelligence Scale

Instructions:

For the following task, you will be presented with a series of drawings contained within a row of boxes. The last box in the series will be empty with dotted lines around the border. The row of boxes to the right of the sample are the answer choices: One of these correctly completes the series. Look at the example below. Answer choice "E" correctly completes the series.

You will have 3 minutes for this task, with a clock counting down at the bottom of the screen. After three minutes elapse, you will automatically move on to the next part. Please click the arrow to start.
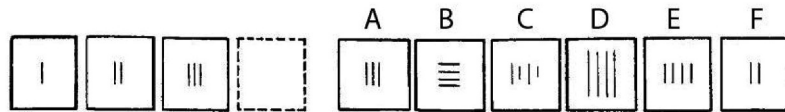
Appendix B: Hierarchical Omega, Neo-FFI Factors

Factor	Omega
Openness	0.866
Conscientiousness	0.934
Extraversion	0.884
Agreeableness	0.821
Neuroticism	0.899

Appendix C: Pearson's correlations for all DVs

Variable	Painting	TCIA_O	TCIA_V	TCIA_T	AUT	Gf	FF	Growth_MS	Fixed_MS	Open	Consc	Extra	Agree	Neurot	VisArt_Act	Activ	Achiev
1. Painting	—																
2. TCIA_O	0.281**	—															
3. TCIA_V	0.175	0.702***	—														
4. TCIA_T	0.202*	0.807***	0.751***	—													
5. AUT	0.077	0.158	0.232*	0.165	—												
6. Gf	0.053	0.338***	0.322***	0.223**	0.293**	—											
7. FF	−0.083	0.131	0.044	−0.004	0.068	0.056	—										
8. Growth_MS	−0.018	0.191	0.165	0.096	0.066	0.074	0.053	—									
9. Fixed_MS	−0.122	−0.168	−0.122	−0.124	−0.180	−0.080	−0.099	−0.508***	—								
10. Open	0.126	0.195	0.148	0.114	0.133	−0.004	0.109	0.257*	−0.297**	—							
11. Consc	−0.195	−0.022	0.036	0.060	−0.041	−0.055	−0.069	0.263*	−0.108	0.010	—						
12. Extra	−0.188	0.093	0.120	0.096	−0.039	0.099	−0.048	0.181	0.074	0.210*	0.374***	—					
13. Agree	0.039	−0.127	0.019	−0.048	0.135	0.055	−0.090	0.315***	−0.414***	0.090	0.177	−0.010	—				
14. Neurot	0.005	−0.076	−0.082	−0.046	−0.029	−0.151	0.039	−0.208*	0.052	0.163	−0.468***	−0.285**	−0.072	—			
15. VisArt_Act	0.112	0.208*	0.241*	0.190	0.020*	0.126	0.043	0.105	−0.082	0.245*	0.017	0.075	−0.056	0.170	—		
16. Activ	0.060	0.307**	0.260**	0.217*	0.021	0.088	−0.010	0.216*	−0.096	0.451***	0.192	0.324***	0.004	0.022	0.736***	—	
17. Achiev	0.015	0.354***	0.312**	0.242*	0.130	0.151	0.113	0.209*	−0.214*	0.318***	0.037	−0.057	0.039	0.150	0.359***	0.486***	—

$* \ p < .05, ** \ p < .005, *** \ p < .001.$

Appendix D: Distribution of Raw Painting Scores

Painting Distribution